HINDI SPEECH CORPORA: A REVIEW

Nivedita, P. Ahmed
School of Engineering and Technology
Sharda University
Greater NODIA, India
nivedita134@gmail.com

Amita Dev
Bhai Parmananda Institute of Business Studies
Govt. of NCT Delhi
New Delhi, India

S. S. Agrawal
KIIT College of Engineering
Gurgaon, India

Abstract* — *A benchmark dataset provides insight into the phenomena that generate the data. Hence, it is an essential requirement to conduct research that requires concept discovery from data. In this paper, we examine the current status of 26 (twenty-six) datasets for Hindi speech (or Hindi speech corpora). This paper also aims at studying their impacts on Hindi speech based computer mediated application development. During this study, we discovered that researchers have paid little attention to issues relating to data collection from a realistic environment through mobile phone. Out of the twenty-six Hindi speech corpora reviewed only one is created for speaker recognition, in which conversation speech samples are recorded through mobile phone for noisy as well as clear condition. .

Keywords—speech corpora, recording enviornment

I. INTRODUCTION

Speech is one of the most important means of communications. Human beings have always been fascinated with the ideas of developing machines that understand and communicate like themselves, but the development of such a machine is a colossal task. The development does not only require creation of hardware components that imitate the auditory and speech production organs, but also demands the development of theories and methods to emulate the complete speech production and perception processes. The latter is the focus of this research that requires ground truth discovery from data to comprehend the underlying phenomena that has generated the speech data. For developing the aforesaid system the prerequisite is a speech corpus. This paper critically examines twenty-six Hindi speech corpora which are readily available, so as to understand the existing work in the field and assess the future path for creating a complete speech corpora which will act as a benchmark for future researches.

In Sections II we present a brief description of different available Hindi speech corpora. In section III, we present a comparative analysis of these Hindi speech corpora. While in section IV we describe ongoing projects in various institutions. In section V, we have presented suggestions and directions for future work.

II. DESCRIPTIONS OF HINDI SPEECH CORPORA

This section briefly describes twenty-six Hindi speech corpora included in our study.

A. Database 1

C-DAC (Center for Development of Advance Computing), NOIDA developed speech corpora for Indian Languages (Hindi, Marathi and Punjabi) [1]. It consists of 1,000 phonetically rich sentences and 10,000 most frequent word set. Phonetically rich sentence consist of words having CO3VCO3 (CO – Consonant, V – Vowel, Consonant varies from 1 to 3) type monosyllable [2] e.g. /ʞɸμɸλ/ (Lotus) consists of syllables /ʞɸ/ and /μɸ/ of type CV and /στριτω/ (womanhood) consist syllable /στρι/ of type CCCV [3]. The text covers words related to digits, days, months, time, units, years etc. It also covers different sentences formed from the above mentioned words as well as some news text. The data set also contains 1,000 prosody rich sentences which covers emotions (like anger, joy, and sadness), question type sentences, negative sentences, command giving sentences and exclamation type sentences [3] shown in figure 1.

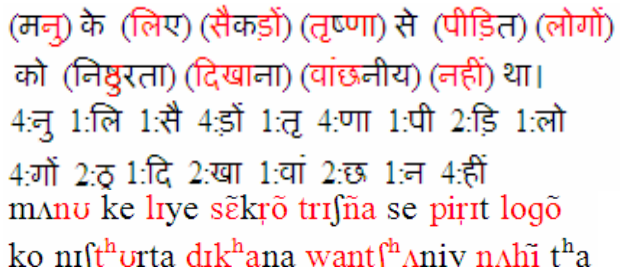

Fig. 1. Example of prosidy rich sentence

Recording has been done by professional speakers (males and females) in noise free and echo canceled condition. They recorded the data at a sampling rate of 44.1 kHz (16 bit) in stereo mode. Tagging of speech units have been done in a hierarchical manner in the order of sentence, words, syllable and phoneme.

B. Database 2 & 3

LILA Hindi L1 database [4] and the LILA Hindi Belt database [5] were created under the LILA project. The goal of LILA project [6] was the collection of speech databases over cellular telephone networks of five languages in three Asian countries. Three languages were recorded in India: Hindi by first language speakers, Hindi by second language speakers and Indian English. Mandarin was recorded in China and Korean in South Korea. These databases belong to SpeechDat-family and follow SpeechDat rules. Both corpora have speech recording of Hindi speakers over the mobile (Global System for Mobile/ Code Division Multiple Access). The speakers belonged to different environments, had diverse demographic profiles, and spoke different dialects. The speech files are stored as sequence of 8-bit, 8 kHz A-law speech files. The database was validated by SPEX (the Netherland) and a SAMPA set was developed for Hindi to provide phonetic representations for the lexicon. Table 1 lists number of speakers according to gender, age and the recording environment.

TABLE I. NUMBER OF SPEAKERS

		Number of Speakers	
		Database 2	***Database 3***
Gender	Males	1012	1011
	Females	1018	1012
Age Group (in years)	16 – 30	965	795
	31 – 45	645	701
	46 – 60	420	527
Mobile User Environment	Passenger in moving car, railway, bus	297	311
	Public Places	299	300
	Stationary pedestrian by road side	305	300
	Home/ Office environment	719	712
	Passenger in moving car using hands – free kit	410	400

Table 2 depicts different items uttered by speakers in databases 2 & 3.

TABLE II. ITEMS UTTERED BY SPEAKERS

Items	**Number of Speakers**	
	Database 2	***Database 3***
Isolated digits	2	2
Sequence of 10 isolated digits	1	2
Connected digits (like telephone numbers, credit card numbers, pin code numbers)	7	7
Natural numbers	1	2
Currency money amount	1	2
Yes/No questions	2	2
Dates (Spontaneous)	3	4
Time phrases	2	3
Application words	6	6
Spotting phrases using an embedded application word	1	2
Directory assistance name	5	7
Spelled words	3	7
Silence words	1	1
Phonetically rich words	4	6
Phonetically rich sentences	13	15
Spontaneous items for control	7	5
Extra Items	Nil	10

C. *Database 4*

EMILLE (Enabling Minority Language Engineering) in collaboration with CIIL (Central Institute of Indian Languages, Mysore) [7, 8] have developed the EMILLE/CIIL Corpus in U.K. It consists of 3 components: monolingual, parallel and annotated. It includes monolingual corpora for 14 South Asian languages (Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Malayalam, Marathi, Oriya, Punjabi, Sinhala, Tamil, Telegu and Urdu). Monolingual Corpora includes written data for all the 14 languages and also contains spoken data for 5 languages viz. Bengali, Guajarati, Punjabi, Urdu and Hindi. It includes 96,157,000 words (93,530,000 written words and 2,627,000 spoken words). The parallel corpora contain 200,000 words of text in English and its accomplished translation in Hindi, Gujarati, Bengali, Urdu and Punjabi. The annotated component includes Urdu and Hindi. It is marked up using CES (Corpus Encoding Standards) – complaint SGML and encoding using Unicode.

D. *Database 5*

Department of Computer Science and Application, Utkal University, Bhubaneswar [9, 10] is developing Text-to-Speech synthesis systems for four Indian languages, namely, Hindi, Odiya, Bengali and Telugu. Text consisted of most of the syllable (Monosyllable, Bisyllable, Trisyllable and some of the Polysyllable) were selected. The speech recording was carried out from native speakers (all males) in laboratory environment; using noise cancellation microphone at a sampling rate of 16 kHz (16 bit).It was observed that syllable level concatenation has smoothness in the utterance of the syllable as compared to the phone level concatenation which leads to the complexity roughness of sound utterance.

E. *Database 6*

TIFR (Tata Institute of Fundamental Research) Mumbai and CDAC NOIDA have developed multi-speaker, continuous speech corpora for Hindi language. It is a comprehensive corpus as it captures phonetic, acoustic, intra-speaker and inter-speaker variability's in Hindi speech. It consists of sets of 10 phonetically rich Hindi sentences spoken by 100 native speakers. Recording was done simultaneously at sampling rate of 16 kHz (16 bit) using two microphones: one positioned closer to speaker and another farther away in a noise free environment. Each speaker was asked to read 10 sentences, each consisting of 2 parts. The first part consists of two 'dialect' sentences which covers the maximum phonemes of Hindi language. The two dialect sentences [11] are shown in figure 2:

धोबिन जब सोकर उठती तो देखती कि चौका साफ पड़ा है और बर्तन मँजे हुए है।

यहाँ से लगभग पाँच मील दक्षिण पश्चिम में कटघर गाँव है।

Fig. 2. Two dialect sentecnes

The second part consists of eight sentences which covers as many pair of broad acoustic classes of phonemes as possible. Every speaker spoke the same first part while the second part was different for every speaker [9, 11].

F. *Database 7*

Indian Institute of Technology (IIT), Kharagpur developed a general purpose speech database for Hindi, Telugu, Tamil, and Kannada languages [9]. The speech samples have been recorded from news bulletin broadcasts. Approximately 3.5 hours of Hindi speech samples were collected from 19 speakers (6 males and 13 females) in a studio in a noise free environment. This database is used for the development of prosody model for speech recognition and speech synthesis, speaker recognition and language identification applications.

G. *Database 8*

IIT Guwahati has prepared a multi-variability speaker recognition database (IITG-MV) [12, 13]. This database has been created to support and evaluate automatic speaker recognition system which is independent of language, environment, channel and style. This

corpus can be used to conduct multiple speech related researches like effect of noise and reverberation, sensor mismatch, speaker diarization etc. Data was collected in four different phases.

In Phase I, they recorded reading and conversation data in English and one favorite language (details of favorite language given in table 4) from 100 speakers in the age group 20 – 40 years over two session. The recording was done in an office environment using 5 different devices viz. headset, microphone, digital voice recorder and mobile in offline mode at a sampling rate of 16 kHz except for mobile on which recording is done at a rate of 8 kHz.

TABLE III. DETAILS OF FAVORITE LANGUAGE IN IITG-MV DATABASE

Languages	Occurrence	
	Phase 1	*Phase 2*
Hindi	28	33
Telugu	10	21
Malayalam	15	8
Oriya	12	11
Bengali	4	9
Assamese	9	6
Gujarati	2	1
Tamil	8	4
Kannada	7	5
Nepali	1	1
Mizo	1	-
Marathi	2	-
English	1	1

In phase II they recorded the data in Hindi, English and one favorite language in reading and conversational style in an uncontrolled environment such as laboratory, hostel rooms, corridors etc. This data was recorded from 100 speakers using the same devices which were used in phase I.

In phase III, speech data were recorded while conversing in all kinds of practical environment like coffee shops, working places, rooms etc at a sampling rate of 8 kHz using different mobile handsets in English and one favorite language from 200 speakers.

In phase IV, they collected the data from speakers coming from all over the India using an Interactive Voice Response (IVR) System. This data was collected in three parts.

- Part-I data involved speech data collection from 55 subjects all over India and to be used for the development of Universal Background Model (UBM).
- Part-II data involved speech data collection from 89 subjects all over India and to be used for the development of speaker models.
- Part-III data is from 197 genuine and 130 imposter trails.

The study showed that there is a significant degradation in performance in case of mismatch in sensor, style and environment in training and testing condition as compared to the matched cases.

H. ***Database 9***

KIIT college of Engineering, Gurgaon with the support of Nokia Research Centre, China developed a text and speech corpus for Hindi and Indian English in the personal communication domain (i.e. message composed on mobile phones and other mobile devices) from 12 different domains (namely vacation report, change of plans, family communication, invitation, congratulations, travel plan, business, feedback, teenagers, school and other domains).The text was collected in raw form which included slangs and then grammatically incorrect data was cleaned using language grammar rules. Then this data was tagged and elaborated to explain context specific meaning of the words. An example [14] of the collected data is shown figure 3. :

Raw	जयकुमार 1000 रु लेकर 10मार्च12 को दूसरी बस से लगभग 3:30 PM में दिल्ली के Hospital पहुँचा
Clean	जय कुमार 1000 रुपया लेकर 10 मार्च 2012 को दूसरी बस से लगभग 3:30 PM पर दिल्ली के Hospital पहुँचा।
Expend	FN/जय/FN LN/कुमार/LN C/1000/C रुपये लेकर D/10/D MN/मार्च/MN Y/2012/Y को O/दूसरी/O बस से लगभग <H/3/H HM/30/HM P_M> पर TN/दिल्ली/TN के &Hospital पहुँचा <पूर्ण_विराम>

Fig. 3. Example of data collected

Texts from 1163 participants from Hindi speaking regions and texts from 1405 participants who were English users were taken and then this data was used to create 13 prompt sheets. These prompt sheets contained 630 phonetically rich sentences. Speech sample of 100 speakers (belonging to different age groups, different profession and qualification) was recorded simultaneously using these prompt sheet on 3 devices (mobile phones, Omni–directional microphone and cardioids microphone) at a sampling rate of 16 kHz in office environment. This database will be used for mobile based speech recognition services [14].

I. ***Database 10***

Linguistic Data Consortium for Indian Language (LDC-IL), Mysore [15] has collected speech database in 22 major Indian languages (Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Indian English Bengali, Indian English Gujarati, Indian English Kannada, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Tamil, Telugu, Urdu) . It become as a linguistic repository of all Indian languages in terms of text, speech and lexical corpora. For Hindi language Raw data is collected for 163 hours 25 minutes and 47 seconds, Segmented data is collected for 105 hours 26 minutes and 45 seconds by 434 speakers, Annotated data is collected for 1 hour 1 minute 28 seconds and Pronunciation Dictionaries(studio recording) 45 hours 51 minutes and 32 seconds.

J. ***Database 11***

International Institute of Information Technology (IIIT) Hyderabad developed an IIIT-H Indic speech database for Bengali, Hindi, Kannada, Malayalam, Marathi, Tamil and Telugu [16]. The purpose of developing this database is to have speech and text corpora made available in the public domain, without copyright restrictions for non-commercial and commercial use. Wikipedia article in these languages were used as text. The recording was done by native speakers of each language and of the age group 20 – 30 years. The recording was done in a professional recording studio using a standard headset microphone connected to a zoom handy recorder. For Hindi language the speech sample of 1 hour and 12 minutes were recorded, where average duration of each utterance was 4.35 seconds.

K. ***Database 12***

Department of Computer Science and Engineering, Rajiv Gandhi University, Arunachal Pradesh (AP) [17] developed a speech corpus for the evaluation of text independent speaker verification system in multilingual environment. It consists of 4 native languages of AP (Nyishi, Adi, Galo and Apatani) along with Hindi and English.

Speech samples of 200 speakers (50 from each native language) of age group 20 years to 50 years were collected in laboratory at a sampling rate of 16 kHz (16 bit) mono-channel using microphones and digital recorders.

*L. **Database 13***

C-DAC NODIA developed a Hindi speech database for Automatic Speech Recognition system for Travel domain. Speech samples of 30 female speakers of age group 17 – 60 years were collected in noise free environment at a sampling rate of 48 kHz (16 bits) in a stereo mode using two microphones. Approximately 26 hours of speech samples were collected which includes 8,567 sentences consisting 74,807 words. The efficiency of system was measured in terms of word recognition rate. The system was tested on 20 speakers (10 from training corpus, 10 for testing corpus). The word recognition rate achieved for training data is 70.73% and for the test data is 60.66% [9, 18].

*M. **Database 14***

IIIT, Hyderabad developed Speech to Speech Synthesis system for Travel and Emergency services in Telugu, Hindi and English languages. The domain of tourism and emergency services was divided into four different sub domains namely: Local travel (D1), Hotel and restaurant transactions (D2), Tourism (D3), and Emergency services (D4). In Hindi *language* 231 sentences were recorded only for the D4 sub domain. The speech sample was recorded from 15 different speakers using a laptop and a standard microphone in noise free environment [9, 19].

*N. **Database 15and 16***

Central Electronics Engineering Research Institute (CEERI) and TIFR jointly developed 2 Hindi speech databases. The first database consisted of 207 words spoken by 50 speakers (two utterances) for Voice Operated Railway Reservation Enquiry Systems. The second database included 1000 phonetically rich sentences spoken by 100 speakers for the development of phoneme based speech recognition systems [2].

*O. **Database 17***

Central Forensic Science Laboratory (CFSL), Chandigarh has developed Speaker Identification Database (SPID) for English and Hindi languages for the development of text independent speaker identification system for forensic applications. Contemporary and Non – Contemporary recording of isolated and contextual speech samples in 10 different modes was done [2].

*P. **Database 18***

ICS, Hyderabad developed speech database for Telugu and Hindi for the development of Interactive Voice Response Systems. They have used the same for synthesizing the content on their website called iKisan.com [2].

*Q. **Database 19***

CFSL, Chandigarh developed multilingual speech database for forensic application. The database consists of speech samples from 10 different languages (Hindi, English, Punjabi, Kashmiri, Urdu, Assamese, Bengali, Tamil, Telugu, and Kannada) from 3 different linguistic groups namely Indo-Aryan, Dravidian and Tibeto-Burman. The data was recorded through 12 different devices (7 microphones, 3 recorders, a mobile phone and a landline telephone) in a room. The data was collected from 100 speakers. Every speaker was to give sample in 3 languages (English, Hindi and a native language). The number of native speaker for each language was 10. This database has been used for Language Independent Speaker Identification System (LISIS) developed by CFSL and Center for Artificial Intelligence and Robotics (CAIR), Bengaluru [4, 20].

*R. **Database 20***

At Devi Ahilya University, Indore speech database has been developed that consists of 1000 Hindi phonetically rich sentences spoken by 100 native speakers. Recording was carried out at a sampling rate of 16 kHz 16-bit Pulse Code Modulation format. This data is being used for studying the effect of noise enhancement techniques for speech recognition and for comparing differences in spoken Hindi in Hindi speaking states and non Hindi speaking states of India [2].

*S. **Database 21***

KIIT, Gurgaon developed an Emotional Speech Database for Hindi language to recognize human emotions [21]. They recorded the data at a sampling rate of 16 kHz (16 bit) in mono channel using condenser microphone in office environment. The database specification details are given in table IV. Emotion recognition experiment has been performed using both short segments of vowels extracted from sentences and Hindi digits, Results showed that both segmental and super segmental parameters were important for emotion recognition.

TABLE IV. DETAILS OF EMOTIONAL SPEECH DATABASE DEVELOPED BY KIIT GURGAON

Database Specification	**Statistics**	
	Emotion Recognition for Isolated word	***Emotion Recognition for Vowel Segment***
Number of Speakers	50	5
Size of Vocabulary	10 Digits	6 Sentences
Repetitions	-	4 Times
Emotions	Neutral, Sad, Surprise, Happy, Fear, Anger	Neutral, Sad, Happy, Fear, Anger
Size of Training Database	3000 Utterances	600 Utterances

*T. **Database 22and 23***

Guru Gobind Singh Indraprastha (GGSIP) University, Delhi developed two Emotional speech databases [22, 23]. For both the databases they recording 5 different emotions namely neutral, anger, happy, surprise and sad. This recording in both the databases was done at a sampling rate of 44.1 kHz (16 bit) mono channel. In the first database 200 Hindi sentences were uttered by 2 speakers (one male and one female). While in the second database 20 Hindi sentence uttered by 10 speakers. The accuracy of perception test on transformed emotion was found out to be 93.2% for surprise, 91.6% for sadness, 83% for happiness and 95.3% for anger [23].

*U. **Database 24***

CDAC NODIA and DRDO (Defence Research & Development Organization) [24] developed Annotated Speech Corpora for three Indian languages namely Hindi, Bengali and Manipuri. Recording was done in multiple environments like normal office, noisy office, and in moving car environment by 30 speakers (both males and females) for 30 minutes at a sampling rate of 16 kHz (16 bit) in mono mode. Manipuri speaker recorded Hindi text for 10 minutes. The recorded text includes any defence related news, 1000 most frequently used words, names of digits, time, days, months, years, units. Annotation of speech units in a hierarchical manner consists of sentence, word, and phoneme. It has been used for development of speech recognition, language identification & speaker identification for defence purpose.

V. *Database 25*

CFSL, Chandigarh developed a multilingual speaker database for forensic applications in Hindi, English, Punjabi and Kannada language. The database consisted of 5,760 samples. These samples were recorded by 80 speakers where each speaker provided the samples in English, Hindi and a native language (40 native speakers of Punjabi and 40 native speakers of Kannada). The recording was done using 6 recording devices and digitized at a sampling rate of 22 kHz (16 bit) [2].

W. *Database 26*

Basic Travel Expression Corpus (BTEC) was developed by NICT, Japan to include travel conversation topics and their translation. It covers Japanese travel expression with English counterpart. The English transcript of this corpus is translated into Hindi language manually in Hindi BTEC corpus [25] under U-STAR (Universal Speech Translation Advance Research) project. The corpus consisted of approx 2k Hindi expression. These speech samples were recorded by Hindi native speakers of age group 17 to 60 years. The recording was carried out a sampling rate of 48 kHz, stereo 16 bit format. For training purpose speech samples were down sampled at 16 kHz, single channel.

III. COMPARISON

The detailed comparison of twenty-five Hindi Speech Corpora (excluding Database 10 because it is general purpose database) is given below.

A. *Database purpose*

Hindi speech corpus is collected for a variety of purposes. The detail classification is given in table V.

TABLE V. DATABASE CLASSIFICATION ON THE BASIS OF PURPOSE

Purposes	*Databases*	*Total*
Speaker Identification/ Speaker Recognition	6, 7, 8, 12, 17, 19, 24, 25	8
Speech Synthesis	1, 5, 7, 11, 14	5
Language Identification	6, 7, 24	3
Emotion Analysis	21, 22, 23	3
Speech Recognition	2, 3, 6, 7, 9, 11, 13, 14, 15, 16, 18, 20, 24	13
Text Translation	4,26	1

In the analyzed set of 25 data sets in this paper, we have found that only 8 out of the given 25 (32%) data sets are used for speaker recognition. Out of these 8, 4 have been developed for forensic and defense purposes, while the remaining 4 have been developed for general purpose use.

B. *Device used for recording*

Table VI below lists down the different devices used to record the data sets. It also mentions the numbers of data sets using these devices.

TABLE VI. DEVICES USED FOR RECORDING

Devices	*Databases*	*Total*
Mobile	2, 3, 8, 9, 19	5
Microphone	5, 6, 7, 8, 9, 11, 12, 13, 14, 19, 21, 22, 23	13
Recorder	8, 17, 19, 25	4
Laptop/Computer	14	1
Landline Telephone	19	1
Not Known	1, 4, 26, 15, 16, 18, 20, 24	8

Out of the 5 data sets which have been recorded on mobiles, 2 are being used for data recognition purpose. These 2 databases use multiple recording devices for developing the database.

C. *Environment*

Table VII below lists various databases according to the environment in which they have been recorded.

TABLE VII. DIFFERENT ENVIRONMENT FOR RECORDING

Environment	*Databases*	*Total*
Noise Free (Professional Labs)	1, 6, 7, 11, 13, 14,	6
Noisy (Hostel Rooms, Office, Moving Car, Corridors etc.)	2, 3, 8, 12, 22, 23, 24, 25	8
Laboratory	5, 8, 9, 17, 19, 21, 24	7
Not Known	4, 26, 15, 16, 18, 20,	6

There is only 1 database which has been recorded in Noisy as well as clean environment. This enables us to make a good training sample and also we can check the variations for testing the model.

D. *Type of text*

Table VIII below lists various types of text being used while recording the databases.

TABLE VIII. DIFFERENT TYPE OF TEXT USED FOR RECORDING

Purposes	*Databases*	*Total*
Running Text	4, 5, 7, 8, 11, 17, 19, 24, 25	9
Sentences	1, 2, 3, 6, 9, 12, 13, 14, 16, 20, 21, 22, 23,26	13
Words	1, 2, 3, 15, 17, 19, 21, 24, 25,26	9
Conversation	8	1
Not Known	18	2

Only 1 database has been recorded using the conversation type text.

On the basis of these comparative studies a Venn diagram is drawn as shown in figure 4. From this we analyzed that only in Database 8 a conversational speech sample was recorded through mobile phone in realistic condition for speaker recognition.

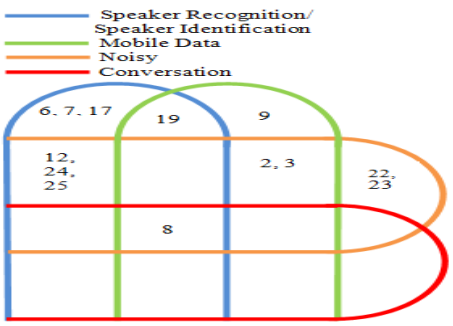

Fig. 4. Venn diagram for database recorded for speaker recognition /speaker identification, using mobile conversation in noisy environment

IV. ONGOING PROJECTS AT VARIOUS INSTITUTIONS

- U-STAR (Universal Speech Translation Advanced Research), an extension of A-STAR project, C-DAC, NOIDA [26]. Its objective is to initiate universal speech to speech translation service among participating countries (Japan, China, Korea, Indonesia, Thailand, India, Vietnam , Singapore, Nepal, Sri Lanka, Mongolia, Bhutan Philippines, Hungary, Germany, Poland, Portugal, France, Turkey, England).
- LDC-IL are developing raw speech corpora for minor languages and extending it for major languages, Domain Specific speech corpora, POS tagged corpora, Chunked corpora, semantically tagged corpora, Syntactic tree bank, and Parallel aligned corpora [27]. It helps in comparing and contrasting structure and functioning of Indian Languages.
- Speech-to-Speech MAT based Dialogue System from Hindi to Indian Languages (Hindi-English, Hindi-Bangla, Hindi-Punjabi, and Hindi-Malyalam) for education and tourism domain [28].

V. SUGGESTIONS AND DIRECTIONS FOR FUTURE WORK

From the comparative study of these databases, we conclude that while recording a single database, there is a need of:

- Using multiple environments
- Using multiple sessions for recording
- Using variety of devices (mobiles, microphones, laptops)
- Using conversation type text
- Using people from different demographic and geographical backgrounds
- Using people from different age group
- Using male and females in a ratio of 1:1

Our research shows that out of the studied 26 databases only 5 are available for research and out of these only 2 are available free of cost. We would like to recommend that the databases created should be readily available and at minimal rates.

VI. REFERENCES

[1] http://www.cdacnoida.in/snlp/stechnology/speech_corpora.asp

[2] S.S.Agarwal, K.Samudravijaya and K.Arora , "Recent advances of speech database development activities for Indian languages", International Symposium on Chinese Spoken Language Processing (ISCSLP 2006), 2006.

[3] K. Arora, S. Arora, K. Verma and S.S. Agrawal, "Automatic extraction of phonetically rich sentences from large text corpus of Indian languages", INTERSPEECH2004 – ICSLP, Jeju, Korea

[4] http://catalog.elra.info/product_info.php?products_id=1071

[5] http://catalog.elra.info/product_info.php?products_id=1170

[6] Eric Sanders, et al., " LILA: Cellular telephone speech database from Asia", in proceeding of LREC, May, 2008

[7] http://www.elda.org/catalogue/en/text/W0037.html

[8] Zhonghua Xiao, Tony McEnery, Paul Baker and Andrew Hardie, "Developing Asian language corpora: Standards and practice", The 4th Workshop on Asian Language Resources, March 2004, Sanya, China.

[9] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, Vishal B. Waghmare. "Indian language speech database: A review", International Journal of Computer Applications, Volume 47 – Number 5 pp. 17 – 21, June, 2012.

[10] Sanghamitra Mohanty, "Syllable based Indian language text to speech system", International Journal of Advances in Engineering & Technology, Vol. 1,Issue 2,pp.138-143,May 2011".

[11] Samudravijaya K, P. V. S. Rao and S. S. Agrawal , "Hindi speech database", Procd. Int. Conf. on Spoken Language processing (ICSLP00), Beijing, China, October 2000.

[12] http://www.iitg.ac.in/eee/emstlab/SRdatabase/introduction.php

[13] Haris B.C. et al., "Multi-variability speech database for robust speaker recognition," National Conference on , Communications (NCC), 2011 vol., no., pp.1-5, 28-30 Jan. 2011

[14] Shyam Agrawal, Shewta Sniha, Pooja Singh and Jesper Olsen, "Development of text and speech database for Hindi and Indian English specific to mobile communication environment", In proceeding of International Conference on The Language Resources and Evaluation Conference, LREC, Istanbul, Turkey,pp. 3415-3421, 2012.

[15] http://www.ldcil.org/resourcesSpeechCorp.aspx

[16] Kishore Prahallad, "The IIIT-H Indic speech database", in proceeding of Interspeech, Portland, USA, September, 2012.

[17] Utpal Bhattacharjee and Kshirod Sarmah, "Development of speech corpus of speaker verification research in multilingual environment", International Journal of Soft Computing and Engineering, Vol. 2, Issue 6,pp. 443-446, January, 2013.

[18] Sunita Arora, Babita Saxena, Karunesh Arora, S.S. Agarwal, "Hindi ASR for travel domain", in proceedings of OCOCOSDA, Kathmandu, Nepal, November,2010.

[19] Anandaswarup V, et al., "Rapid development of speech to speech systems for tourism and emergency services in Indian languages", In proceeding of International Conference on services in Emergency Markets, Hyderabad, India, 2010.

[20] Shivani Sharma, S.K. Jain, R.M. Sharma and S.S. Agrawal, "Present scenario of forensic speaker identification in India", In proceeding of O-COCOSDA, Nepal, 2010.

[21] Agarwal S. S., "Emotions in Hindi speech-analysis, perception and recognition", International conference on Speech Database and Assessment (Oriental COCOSDA), Taiwan, pp. 7-13, October 2011.

[22] Maninder Singh Suri, Divya Setia and Anurag Jain, "Praat implementation for prosody conversion", In proceeding of the 4th National Conference; INDIACom-2010, Feburary, 2010.

[23] Agrawal S.S., Nupur Prakash, Anurag Jain, "Transformation of emotion based on acoustic features of intonation patterns for Hindi speech", IJCSNS International Journal of Computer Science and Network Security, Vol.10 No.9, pp. 198-205, September 2010.

[24] http://www.cdacnoida.in/snlp/stechnology/corpora_DRDO.asp

[25] Sunita Arora, Karunesh Arora, S.S. Agarwal, " Statistical analysis of Hindi BTEC speech database", in International Conference on Speech Database and Assessment (Oriental COCOSDA) ,Macau, 2012.

[26] http://www.cdacnoida.in/snlp/ongoing_projects/ustar.asp

[27] http://www.ldcil.org/areasOfWorkCorpCreat.aspx

[28] http://www.cdackolkata.in/Groups.php?gid=IT_Systems&lang=English